Statistical Inference Course Project (Part 2)

BELMOUIDI Mohamed

28 December, 2024

Contents

Instructions

- Load the ToothGrowth data and perform basic exploratory analyses.
- Summarize the data.
- Use confidence intervals and/or hypothesis tests to compare tooth growth by supplement and dose.
- State conclusions and assumptions.

Exploratory Data Analysis

```
library(dplyr, warn.conflicts = F)
library(ggplot2)
library(ggthemes)

ToothGrowth <- tbl_df(ToothGrowth)
```

Data Overview

```
## Warning: `tbl_df()` was deprecated in dplyr 1.0.0.
## i Please use `tibble::as_tibble()` instead.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was generated.
```

```
ToothGrowth %>% str()
```

```
## tibble [60 x 3] (S3: tbl_df/tbl/data.frame)
##  $ len : num [1:60] 4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

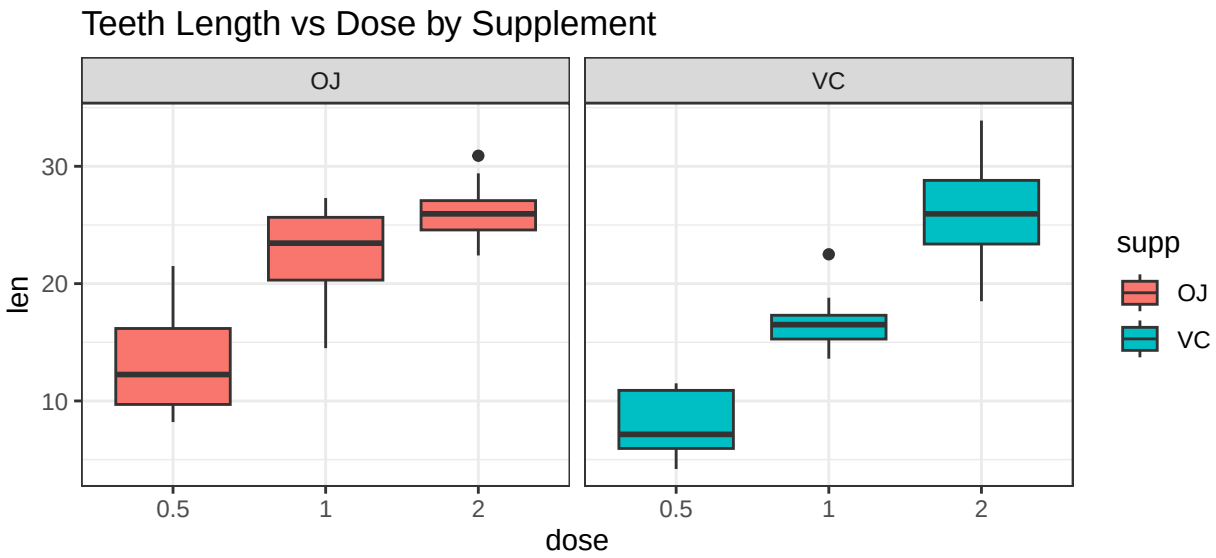
```r
ToothGrowth %>% summary()
```

```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

The dataset has 60 observations and 3 variables:

- **len**: Tooth length (numeric).
- **supp**: Supplement type (factor, VC or OJ).
- **dose**: Dose in mg (numeric, converted to factor).

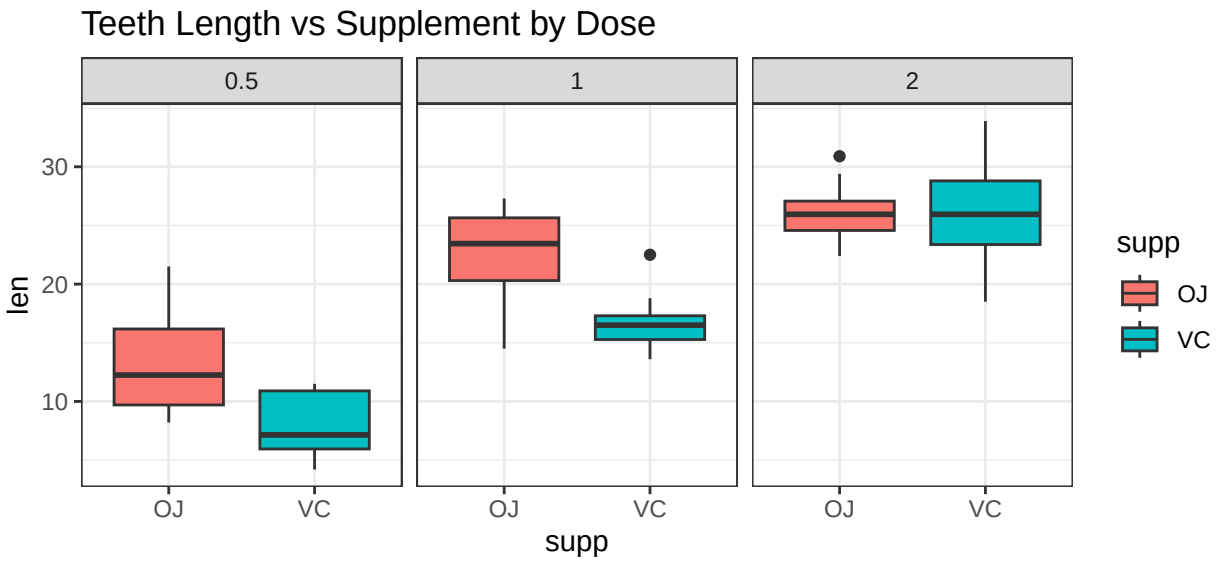
```r
ToothGrowth %>%
  ggplot(aes(x = dose, y = len, fill = supp)) +
  geom_boxplot() +
  facet_grid(. ~ supp) +
  theme_bw() +
  ggtitle("Teeth Length vs Dose by Supplement")
```



Plots

Teeth length increases with dose for both supplement types.

```r
ToothGrowth %>%
  ggplot(aes(x = supp, y = len)) +
  geom_boxplot(aes(fill = supp)) +
  facet_wrap(~ dose) +
  theme_bw() +
  ggtitle("Teeth Length vs Supplement by Dose")
```



At doses 0.5 and 1, OJ is more effective, while VC slightly outperforms at dose 2.

Hypothesis Tests

By Dose Levels Dose 0.5 vs 1

```
t.test(len ~ dose, data = ToothGrowth %>% filter(dose %in% c(0.5, 1)))
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means between group 0.5 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

Dose 0.5 vs 2

```
t.test(len ~ dose, data = ToothGrowth %>% filter(dose %in% c(0.5, 2)))
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means between group 0.5 and group 2 is not equal to 0
## 95 percent confidence interval:
```

```
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

Dose 1 vs 2

```
t.test(len ~ dose, data = ToothGrowth %>% filter(dose %in% c(1, 2)))
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 0.00001906
## alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

In all cases, p-values are < 0.05, indicating significant differences in teeth length by dose.

```
t.test(len ~ supp, data = ToothGrowth)
```

By Supplement Type

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means between group OJ and group VC is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

P-value is 0.06, indicating no significant difference in teeth length between supplement types.

Conclusions

- **Dose Levels**: Significant positive relationship between dose and teeth length.
- **Supplement Types**: No statistically significant difference; OJ is more effective at lower doses, while VC slightly outperforms at dose 2.

Assumptions

- **Independence**: Random sampling/assignment.
- **Normality**: Population distributions are normal or quasi-normal.

Assuming these conditions are met, the results are valid.